Price decomposition in large-scale stochastic optimal control

Kengy Barty, Pierre Carpentier, Guy Cohen, Pierre Girardeau

PRICE DECOMPOSITION IN LARGE-SCALE STOCHASTIC OPTIMAL CONTROL

KENGY BARTY, PIERRE CARPENTIER, GUY COHEN, AND PIERRE GIRARDEAU

Abstract. We are interested in optimally driving a dynamical system that can be influenced by exogenous noises. This is generally called a Stochastic Optimal Control (SOC) problem and the Dynamic Programming (DP) principle is the natural way of solving it. Unfortunately, DP faces the so-called curse of dimensionality: the complexity of solving DP equations grows exponentially with the dimension of the information variable that is sufficient to take optimal decisions (the state variable).

For a large class of SOC problems, which includes important practical problems, we propose an original way of obtaining strategies to drive the system. The algorithm we introduce is based on Lagrangian relaxation, of which the application to decomposition is well-known in the deterministic framework. However, its application to such closed-loop problems is not straightforward and an additional statistical approximation concerning the dual process is needed. We give a convergence proof, that derives directly from classical results concerning duality in optimization, and enlghten the error made by our approximation. Numerical results are also provided, on a large-scale SOC problem. This idea extends the original DADP algorithm that was presented by Barty, Carpentier, and Girardeau (2010).

Introduction

Consider a controlled dynamical system over a discrete and finite time horizon. This system may be influenced by exogenous noises that affect its behaviour. We suppose that, at every instant, the decision maker is able to observe these noises and to keep these observations in memory. Since it is generally profitable to take available observations into account when designing future decisions, we are looking for strategies rather than simple decisions. Such strategies (or policies) are feedback functions that map every instant and every possible history of the system to a decision to be made.

More precisely, we are here interested in optimization problems with a large number of variables. The typical application we have in mind is the following. Consider a power producer that owns a certain number of power units. Each unit has its own local characteristics such as physical constraints that restrain the set of feasible decisions, and production costs that depend on the type of fuel that is used to produce power. The power producer has to control the power units so that a global power demand is met at every instant. The power demand, as well as other parameters such as inflows in water reservoirs or unit breakdowns, are random. Naturally, he is looking for strategies that make the production cost minimal, over a given time horizon. In such a problem, both the number of power units and the number of time steps are usually large.

Date: December 15, 2010.

1991 *Mathematics Subject Classification.* 93E20, 49M27, 49L20.

Key words and phrases. Stochastic optimal control, Decomposition methods, Dynamic Programming.

One classical approach when dealing with stochastic dynamic optimization problems is to discretize the random inputs of the problem using scenario trees. Such an approach has been widely studied within the Stochastic Programming community (see the book by Shapiro, Dentcheva, and Ruszczyński, 2009, for an overview of this methodology). One of the advantages of such a technique is that as soon as the scenario tree is drawn, the derived problem can be treated by classical Mathematical Programming techniques. Thus, a number of decomposition methodologies have been proposed (Higle and Sen, 1996, Carpentier, Cohen, Culioli, and Renaud, 1996, Ruszczyński and Shapiro, 2003, Chapter 3) and even applied to energy planning problems (Bacaud, Lemaréchal, Renaud, and Sagastizábal, 2001). A general theoteric point of view concerning the way to combine the discretization of expectation together with the discretization of information is given by Barty (2004). However, in a multi-stage setting, this methodology suffers from the drawbacks that arise with scenario trees. As it was pointed out by Shapiro (2006), the number of scenarios needed to achieve a given accuracy grows exponentially with the number of time steps of the problem.

The other natural approach to solve SOC problems is to rely on the Dynamic Programming (DP) principle (see Bellman, 1957, Bertsekas, 2000). The core of the DP approach is the definition of a state variable that is, roughly speaking, the variable that, in conjunction with the time variable, is sufficient to take an optimal decision at every instant. It does not have the drawback of the scenario trees concerning the number of time steps since strategies are, in this context, depending on a state variable whose space dimension usually does not grow with time[1]. However, DP suffers from another drawback which is the so-called *curse of dimensionality*: the complexity of solving the DP equation grows exponentially with the state space dimension. Hence, brutally solving the DP equation is generally intractable when the state space dimension goes beyond several units. Recently, Vezolle, Vialle, and Warin (2009) were able to solve it on a 10-state-variables energy management problem, using parallel computation coupled with adequate data distribution.

Another popular idea is to represent the value functions (solutions of the DP equation) as a linear combination of a priori chosen basis functions (see among others Bellman and Dreyfus, 1959, Bertsekas and Tsitsiklis, 1996, Sect. 6.5). This approach, called Approximate Dynamic Programming or often Least-Squares Monte-Carlo, has also become very popular in the context of American option pricing through the work of Longstaff and Schwartz (2001). This approximation reduces the complexity of solving the DP equation drastically. However, in order to be practically efficient, such an approach requires some a priori information about the problem, in order to define a well suited functional subspace. Indeed, there is no systematic means to choose the basis functions and several choices have been proposed in the literature (de Farias and Van Roy, 2003, Tsitsiklis and Van Roy, 1996, Bouchard and Warin, 2010).

When dealing with large-scale optimization problems, the decomposition/coordination approach aims at finding a solution to the original problem by iteratively solving smaller-dimensional subproblems. In the deterministic case, several types of decomposition have been proposed (e.g. by prices or by quantities) and unified in a general framework using the Auxiliary Problem Principle by Cohen (1980a). In the open-loop stochastic case, i.e. when controls do not rely on any observation, Cohen and Culioli (1990) proposed to take advantage of both decomposition techniques and stochastic gradient algorithms. These techniques have been extended in the closed-loop stochastic case by Barty, Roy, and Strugarek (2009), but so far

[1]In the case of power management, the state dimension is usually the number of power units.

they fail to provide decomposed state dependent strategies in the Markovian case. This is because a subproblem optimal strategy depends on the state of the whole system, not only on the local state. In other words, decomposition approaches are meant to decompose the control space, namely the range of the strategy, but the numerical complexity of the problems we consider here also arises because of the dimensionality of the state space, that is to say the domain of the strategy.

We here propose a way to use price decomposition within the closed-loop stochastic case. The coupling constraints, namely the constraints preventing the problem from being naturally decomposed, are dualized using a Lagrange multiplier (price). At each iteration, the price decomposition algorithm solves each subproblem using the current price, then uses the solutions to update the price. In the stochastic context, price is a random process whose dynamics is not available, so the subproblems do not in general fall into the Markovian setting. However, in a specific instance of this problem, Strugarek (2006) exhibited a dynamics for the optimal multiplier, and he showed that these dynamics were independent with respect to the decision variables. Hence it was possible to come down to the Markovian framework and to use DP to solve the subproblems in this case. Following this idea, Barty et al. (2010) proposed to choose a parametrized dynamics for these multipliers in such a way that solving subproblems using DP becomes possible. While the approach, called Dual Approximate Dynamic Programming (DADP), showed promising results on numerical examples, it suffers from the fact that the induced restrained dual space is non-convex. This led to some numerical instabilities and, probably more important, it was not possible to give convergence results for the algorithm. We here propose to extend DADP in a more general way that allows us to derive convergence results and solves the problem of numerical instabilities.

The paper is organized as follows. In Section 1, we present the general SOC problem and the DP principle. Then we concentrate on a more specific class of problems, that we call decomposable problems, and recall the previous version of the DADP algorithm. In Section 2, we present the new version we propose and give convergence results for the algorithm. Finally, in Section 3, we apply DADP to two numerical examples, the first being the one from the previous paper by Barty et al. (2010) and the second one being a more realistic power management example.

1. Mathematical formulation

1.1. **General problem setting.** All along the paper, random variables are denoted using **bold** letters. Consider a discrete and finite time horizon $0,\ 1,\ \ldots,\ T$ and a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. To define a stochastic dynamical system, we need:

- a stock process $\boldsymbol{X} = (\boldsymbol{X}_0, \ldots, \boldsymbol{X}_T)$ which represents the physical states of the system through time, the value of $\boldsymbol{X}_t$ lying, at every instant t, in a Hilbert space $\mathbb{X}_t$;
- a control process $\boldsymbol{U} = (\boldsymbol{U}_0, \ldots, \boldsymbol{U}_{T-1})$, the value of $\boldsymbol{U}_t$ lying, at every instant t, in a Hilbert space $\mathbb{U}_t$;
- a noise process $\boldsymbol{W} = (\boldsymbol{W}_0, \ldots, \boldsymbol{W}_{T-1})$, the value of $\boldsymbol{W}_t$ lying, at every instant t, in a Hilbert space $\mathbb{W}_t$.

The spaces $\mathbb{X}_t$, $\mathbb{U}_t$ and $\mathbb{W}_t$ are generally finite-dimensional spaces. In the sequel, we suppose $\mathbb{X}_t = \mathbb{R}^n$ and $\mathbb{U}_t = \mathbb{R}^m$. The decision variable $\boldsymbol{U}_t$ being a random variable, and our purpose being to use variational techniques that require the notion of gradient, it is natural to suppose that $\boldsymbol{U}_t$ lies in a Hilbert space $\mathcal{U}_t$, for example $L^2(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{U}_t)$.

The three types of variables are linked together in the following way. At every time step t, there exists a function f_t (the dynamics of the system) that maps the

triplet $(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_t)$ to the next stock value $\boldsymbol{X}_{t+1}$. Let $(\mathcal{A}_0, \dots, \mathcal{A}_{T-1})$ be the filtration associated with the stochastic process $\boldsymbol{W}$. We suppose that, at every time step t, the decision maker is able to observe and to keep in memory all the past history of $\boldsymbol{W}$ up to time t. The causality principle states that the decision $\boldsymbol{U}_t$ at time t is $\mathcal{A}_t$-measurable, i.e. only depends on past observations. Moreover, at each time step t, a cost $C_t(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_t)$ is incurred. Finally, at the final time T, a cost $K(\boldsymbol{X}_T)$ is added. The Stochastic Optimal Control (SOC) problem we would like to solve hence reads:

$$\min_{\boldsymbol{X},\boldsymbol{U}} \quad \mathbb{E}\left(\sum_{t=0}^{T-1} C_t\left(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_t\right) + K\left(\boldsymbol{X}_T\right)\right), \tag{1a}$$

subject to dynamics constraints:

$$\boldsymbol{X}_{t+1} = f_t\left(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_t\right), \qquad \forall t = 0, \dots, T-1, \tag{1b}$$

$$\boldsymbol{X}_0 \text{ is given}, \tag{1c}$$

as well as bound constraints:

$$\underline{x}_t \leq \boldsymbol{X}_t \leq \overline{x}_t, \qquad \forall t = 1, \dots, T, \tag{1d}$$

$$\underline{u}_t \leq \boldsymbol{U}_t \leq \overline{u}_t, \qquad \forall t = 0, \dots, T-1, \tag{1e}$$

static constraints:

$$g_t\left(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_t\right) = 0, \qquad \forall t = 0, \dots, T-1, \tag{1f}$$

and the non-anticipativity constraint:

$$\boldsymbol{U}_t \text{ is } \mathcal{A}_t\text{-measurable}. \tag{1g}$$

Constraints (1b), (1d), (1e) and (1f) have to be understood in the $\mathbb{P}$-almost sure sense. We give examples for constraint (1f) in §2. With no further assumptions, Problem (1) cannot generally be solved analytically, except for quite particular cases among which is, for instance, the Linear Quadratic Gaussian (LQG) case. One has to be aware that, when solving this problem, one is looking for functions that map every possible history of the system to a decision; the domain of such a function is clearly growing with time and representing it on a computer rapidly becomes intractable.

1.2. **The Dynamic Programming Principle.** Fortunately enough, control theory helps us reduce the size of the optimal strategy's domain in some cases. Let us first make the following assumption.

Assumption 1. Noises $\boldsymbol{W}_0, \dots, \boldsymbol{W}_{T-1}$ are independent over time.

Now define functions V_t, for every time step $t = 0, \dots, T$, as:

$$V_t(x) = \min_{\substack{\boldsymbol{X}_t, \dots, \boldsymbol{X}_T \\ \boldsymbol{U}_t, \dots, \boldsymbol{U}_{T-1}}} \mathbb{E}\left(\sum_{s=t}^{T-1} C_s\left(\boldsymbol{X}_s, \boldsymbol{U}_s, \boldsymbol{W}_s\right) + K\left(\boldsymbol{X}_T\right) \,\middle|\, \boldsymbol{X}_t = x\right), \qquad \forall x \in \mathbb{X}_t,$$

subject to the same[2] constraints as in Problem (1). Function V_t represents the minimal remaining cost of the problem when starting at time t, for every possible stock value x.

Under Assumption 1, the Dynamic Programming (DP) principle states that the variable $\boldsymbol{X}_t$, along with the current noise value $\boldsymbol{W}_t$, contains all the information that is sufficient to take the optimal decision at time t, hence the term *state* variable.

[2]while starting at time t

Moreover, it provides a way to compute functions V_t, that we now call Bellman functions (or value functions), as well as optimal strategy, in a backward manner.

$$V_T(x) = K(x), \qquad \forall x \in \mathbb{X}_T, \tag{2a}$$

and, for every time step $t = T-1, \dots, 0$:

$$V_t(x) = \mathbb{E}\left(\min_u C_t(x, u, \boldsymbol{W}_t) + V_{t+1}(f_t(x, u, \boldsymbol{W}_t))\right), \qquad \forall x \in \mathbb{X}_t. \tag{2b}$$

Compared with the original setting where the optimal strategy domain was growing along with time steps, the DP principle drastically reduces the size of the information needed to make an optimal decision.

Remark 1 (About the overtime independence). In the case when the model is such that noises that affect the system have some sort of correlation through time, one can always explicit the dynamics of the noise variable and add it to the dynamics of $\boldsymbol{X}_t$, thus defining a new (albeit larger!) state variable as well as a new noise variable that is now independent over time.

Remark 2 (Hazard-Decision setting). The reader may have noticed that the way the non-anticipativity constraint in written allows the decision maker at time t to observe the current noise value $\boldsymbol{W}_t$ before choosing the control $\boldsymbol{U}_t$. In such a setting the optimal decision at time t depends on both the state variable $\boldsymbol{X}_t$ and the noise variable $\boldsymbol{W}_t$ whereas the value function only depends on the state variable $\boldsymbol{X}_t$.

Note however that the dimension of the state space $\mathbb{X}_t$ might still be quite large. Yet the complexity of solving the DP equation (2) grows exponentially with the dimension of $\mathbb{X}_t$; this unpleasant feature is well known as the *curse of dimensionality* and prevents us from solving this equation by discretization when the state space dimension is, say, greater than 5.

1.3. **Decomposable problem setting.** Let us now present a particular instance of Problem (1) on which we are able to reduce even more the size of the information needed to take a reasonable decision.

We consider a system which consists of N subsystems[3], whose dynamics and cost functions are independent one from another. More precisely, the state $\boldsymbol{X}_t$ (respectively the control $\boldsymbol{U}_t$) of the global system writes $\left(\boldsymbol{X}_t^1, \dots, \boldsymbol{X}_t^N\right)$ with $\boldsymbol{X}_t^i \in L^2(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R}^{n_i})$ (resp. $\left(\boldsymbol{U}_t^1, \dots, \boldsymbol{U}_t^N\right)$ with $\boldsymbol{U}_t^i \in L^2(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R}^{m_i})$) and $n = \sum_{i=1}^N n_i$ (resp. $m = \sum_{i=1}^N m_i$), so that the global dynamics $\boldsymbol{X}_{t+1} = f_t(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_t)$ can be written independently unit by unit: $\boldsymbol{X}_{t+1}^i = f_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t\right)$, $i = 1, \dots, N$. In the same way, the global cost $C_t(\boldsymbol{X}_t, \boldsymbol{U}_t, \boldsymbol{W}_t)$ is equal to the sum of the local unit costs $C_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t\right)$, $i = 1, \dots, N$. At the end of the time period, each unit i causes a cost K^i that only depends on its final state $\boldsymbol{X}_T^i$.

Remark that, without further constraints, the induced SOC problem can be stated independently unit by unit, though the same noise variable affects all units (see Appendix B for a precise proof). Hence, under Assumption 1, the solving of the DP equation can be decomposed unit by unit. For each unit, the optimal strategy depends only on its local state[4], which is usually far smaller than the dimension of the global state space.

Consider now a static constraint (1f) that couples the units together. We suppose that such a coupling arises from a set of static $\mathbb{R}^d$-valued constraints, the constraint at time step t reading $\sum_{i=1}^N g_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t\right) = 0$. This kind of coupling constraint is natural in many industrial applications, including the case of a power management

[3]We often use the term "*units*" for subsystems.

[4]and on the noise at the current time step because we are in the Hazard-Decision setting

problem that we already mentioned in the introduction: the sum of the productions of the power units must meet an uncertain power demand.

The decomposable problem we are interested in solving in the following reads:

$$\min_{\boldsymbol{X},\boldsymbol{U}} \quad \mathbb{E}\left(\sum_{t=0}^{T-1}\sum_{i=1}^{N} C_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t\right) + \sum_{i=1}^{N} K^i\left(\boldsymbol{X}_T^i\right)\right) \tag{3a}$$

subject to dynamics constraints:

$$\boldsymbol{X}_{t+1}^i = f_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t\right), \qquad \forall t = 0,\dots,T-1, \forall i = 1,\dots,N, \tag{3b}$$

$$\boldsymbol{X}_0^i \text{ is given}, \qquad \forall i = 1,\dots,N, \tag{3c}$$

as well as bound constraints:

$$\underline{x}_t^i \leq \boldsymbol{X}_t^i \leq \overline{x}_t^i, \qquad \forall t = 1,\dots,T, \forall i = 1,\dots,N, \tag{3d}$$

$$\underline{u}_t^i \leq \boldsymbol{U}_t^i \leq \overline{u}_t^i, \qquad \forall t = 0,\dots,T-1, \forall i = 1,\dots,N, \tag{3e}$$

static constraints:

$$\sum_{i=1}^{N} g_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t\right) = 0, \qquad \forall t = 0,\dots,T-1, \tag{3f}$$

and the non-anticipativity constraint:

$$\boldsymbol{U}_t^i \text{ is } \mathcal{A}_t\text{-measurable}, \qquad \forall t = 0,\dots,T-1, \forall i = 1,\dots,N. \tag{3g}$$

There are three types of coupling in Problem (3):

- The first comes from the state dynamics (3b) that induce a temporal coupling.
- The second one arises from the static constraints (3f) that induce a spatial coupling: they link together all the subsystems at each time step t.
- The third type of coupling is informational: it comes from the causality constraint (3g), which prevents us from decomposing directly scenario by scenario : if two realizations of the noise process are identical up to time t, then the same control has to be applied at time t on both realizations.

Constraints (3f) prevent us from decomposing the optimization problem unit by unit: the solution $\boldsymbol{U}_t^i$ for unit i and time t has to be searched as a feedback function φ_t^i depending on the current noise value and on the whole stock variable $\boldsymbol{X}_t = (\boldsymbol{X}_t^1, \dots, \boldsymbol{X}_t^N)$ rather than on the local stock variable $\boldsymbol{X}_t^i$! Adding the coupling constraint (3f) drastically changed the structure of the problem.

Remark 3 (Local and global noises). Applications we have in mind are power management problems which are completely "flower-shaped", in the following sense. The noise variable $\boldsymbol{W}_t$ at time t is composed of two different kinds of noise:

- a *local* noise $\boldsymbol{W}_t^i$ for every subsystem i, i.e. at every petal of the flower (uncertain inflows entering a water reservoir, for instance);
- a *global* noise $\boldsymbol{D}_t$ at the center of the flower (a total power demand, for instance).

In such a setting, only the local noise appears in the cost function and in the dynamics, leading to functions of the form:

$$C_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t^i\right) \text{ and } f_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t^i\right),$$

while the global noise appears only in the coupling constraint as, for instance:

$$\sum_{i=1}^{N} g_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i\right) = \boldsymbol{D}_t.$$

Keeping this particular case in mind shall give us some insight about how to decompose the global problem as well as possible. This is explained in more details in §2.1 and such settings are treated in the numerical experiments of §3.

1.4. **Previous paper.** In a previous study (Barty et al., 2010), the authors proposed a way of handling Problem (3) by approximate Lagrangian decomposition. The proposed algorithm, called Dual Approximate Dynamic Programming (DADP) is as follows. Let us introduce the Lagrangian of Problem (3):

$$\mathcal{L}\left(\boldsymbol{X}, \boldsymbol{U}, \boldsymbol{\lambda}\right) := \mathbb{E}\Bigg(\sum_{t=0}^{T-1} \sum_{i=1}^{N} \left(C_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t\right) + \boldsymbol{\lambda}_t^\top g_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t\right)\right) + \sum_{i=1}^{N} K^i\left(\boldsymbol{X}_T^i\right) \Bigg),$$

with $\boldsymbol{\lambda}_t \in L^2(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R}^d)$ the Lagrange multiplier of the coupling constraint (3f) and $\boldsymbol{\lambda} := (\boldsymbol{\lambda}_0, \ldots, \boldsymbol{\lambda}_{T-1})$. Note that, since the dualized constraint is $\mathcal{A}_t$-measurable, the Lagrange multiplier $\boldsymbol{\lambda}_t$ need only to have the same measurability.

Problem (3) is always equivalent to:

$$\min_{\boldsymbol{X}, \boldsymbol{U}} \max_{\boldsymbol{\lambda}} \quad \mathcal{L}\left(\boldsymbol{X}, \boldsymbol{U}, \boldsymbol{\lambda}\right),$$

where the minimization is subject to all constraints of Problem (3) except constraint (3f). If $\mathcal{L}$ has a saddle point (see Appendix A for a definition and a characterization of saddle points), then this problem is equivalent to the so-called dual problem:

$$\max_{\boldsymbol{\lambda}} \min_{\boldsymbol{X}, \boldsymbol{U}} \quad \mathcal{L}\left(\boldsymbol{X}, \boldsymbol{U}, \boldsymbol{\lambda}\right), \tag{4}$$

under, once again, the same constraints as in Problem (3) except the coupling constraint (3f).

The key point of the so-called price decomposition algorithm is that the inner minimization problem can be split into N subproblems, each one involving a single subsystem (once again, see Appendix B for more details). One might think that solving these subproblems is much simpler than solving the original global problem. This is not the case here: because the dual variable $\boldsymbol{\lambda}$ is a stochastic process that depends in general on the whole history of the system, we cannot reasonably make the overtime independence assumption that leads to the DP principle and subproblems are just as hard as Problem (1)!

The idea of Barty et al. (2010) is to force the dual process to satisfy a prescribed dynamics:

$$\boldsymbol{\lambda}_0 = h_{\alpha_0}\left(\boldsymbol{W}_0\right), \tag{5a}$$

$$\boldsymbol{\lambda}_{t+1} = h_{\alpha_{t+1}}\left(\boldsymbol{\lambda}_t, \boldsymbol{W}_{t+1}\right), \qquad \forall t = 0, \ldots, T-2, \tag{5b}$$

where h_{α_t} is an a priori chosen function parametrized by $\alpha_t \in \mathbb{R}^q$. We note $\alpha = (\alpha_0, \ldots, \alpha_{T-1})$. Given a vector α^k of coefficients at iteration k of the algorithm which defines the current values of the dual variables, the first step of DADP is to solve the N subproblems by DP with state $(\boldsymbol{X}_t^i, \boldsymbol{\lambda}_t)$. In order to update the Lagrange multipliers, the authors propose to draw S trajectory samples of the noise $\boldsymbol{W}$ and integrate the dynamics (3b)–(3c) and (5) using the optimal feedback

laws obtained at the first step, thus obtaining S sample trajectories of $\boldsymbol{X}^k$, $\boldsymbol{U}^k$ and $\boldsymbol{\lambda}^k$. A gradient step is then performed sample by sample:

$$\boldsymbol{\lambda}_t^{k+\frac{1}{2},s} = \boldsymbol{\lambda}_t^{k,s} + \rho_t \times \sum_{i=1}^{N} g_t^i \left(\boldsymbol{X}_t^{i,k,s}, \boldsymbol{U}_t^{i,k,s}, \boldsymbol{W}_t^s \right), \qquad \forall s = 1, \dots, S,$$

with ρ_t obeying the rules of the step-size choice in Uzawa's algorithm (see Appendix A). Finally, we solve the following regression problem:

$$\min_{\alpha_0,\dots,\alpha_{T-1}} \sum_{s=1}^{S} \left(\left\| h_{\alpha_0}\left(\boldsymbol{W}_0^s\right) - \boldsymbol{\lambda}_0^{k+\frac{1}{2},s} \right\|_{\mathbb{R}^d}^2 + \sum_{t=0}^{T-2} \left\| h_{\alpha_{t+1}}\left(\boldsymbol{\lambda}_t^{k+\frac{1}{2},s}, \boldsymbol{W}_{t+1}^s\right) - \boldsymbol{\lambda}_{t+1}^{k+\frac{1}{2},s} \right\|_{\mathbb{R}^d}^2 \right).$$

The last minimization produces coefficients α^{k+1} which define, using Equation (5), a new process $\boldsymbol{\lambda}^{k+1}$.

This procedure has several advantages, notably that its complexity is linear with respect to the number N of subproblems and that it may lead, depending on the choice for the dual dynamics h, to tractable approximations of the original problem. The authors illustrate this fact on a small example on which they are able to compare standard DP and DADP.

Still, it has some drawbacks, mainly theoretical. First of all, the shape of the dynamics introduced for the dual process is arbitrarily and once for all chosen and the quality of the result depends on this choice. Moreover, this dynamics defines a subspace which is non-convex. The next iterate $\boldsymbol{\lambda}^{k+1}$ being a projection on this subspace, it is not well defined and some oscillations observed in practice may be due to this fact. Finally, this non-convexity prevents us from obtaining convergence results for this algorithm.

2. Dual Approximate Dynamic Programming revisited

We now propose a new version of the DADP algorithm and show how it overcomes the above mentioned drawbacks encountered with the original algorithm. In this new approach, we do not suppose a given dynamics for the multipliers anymore. Still, we use the standard price decomposition algorithm and perform the update of the multipliers scenario-wise using the classical gradient step:

$$\boldsymbol{\lambda}_t^{k+1,s} = \boldsymbol{\lambda}_t^{k,s} + \rho_t \times \sum_{i=1}^{N} g_t^i \left(\boldsymbol{X}_t^{i,k,s}, \boldsymbol{U}_t^{i,k,s}, \boldsymbol{W}_t^s \right), \qquad \forall s = 1, \dots, S.$$

The difficulty is now to solve the subproblems, as explained in §2.1.

2.1. Projection of the dual process.

After Lagrangian decomposition of Problem (3) with a given multiplier $\boldsymbol{\lambda}$, the i-th subproblem reads:

$$\min_{\boldsymbol{X}^i, \boldsymbol{U}^i} \quad \mathbb{E}\left(\sum_{t=0}^{T-1} \left(C_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t\right) + \boldsymbol{\lambda}_t^\top g_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t\right) \right) + K^i\left(\boldsymbol{X}_T^i\right) \right) \tag{6a}$$

subject to dynamic constraints:

$$\boldsymbol{X}_{t+1}^i = f_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t\right), \qquad \forall t = 0, \dots, T-1, \tag{6b}$$

$$\boldsymbol{X}_0^i \text{ is given}, \tag{6c}$$

as well as bound constraints:

$$\underline{x}_t^i \leq \boldsymbol{X}_t^i \leq \overline{x}_t^i, \qquad \forall t = 1, \dots, T, \tag{6d}$$

$$\underline{u}_t^i \leq \boldsymbol{U}_t^i \leq \overline{u}_t^i, \qquad \forall t = 0, \dots, T-1, \tag{6e}$$

and the non-anticipativity constraint:

$$\boldsymbol{U}_t^i \text{ is } \mathcal{A}_t\text{-measurable.} \tag{6f}$$

As it was already mentioned, since the dual stochastic process $\boldsymbol{\lambda}$ generally depends on the whole history of the process, solving this problem is in general as complex as solving the original problem. In order to bypass this difficulty, let us choose at each time step t a random variable $\boldsymbol{Y}_t^i$ that is measurable with respect to $\mathcal{A}_t$. We call $\boldsymbol{Y}^i = (\boldsymbol{Y}_0^i, \dots, \boldsymbol{Y}_{T-1}^i)$ the information process for subsystem i. The idea is to rely on a short memory process $\boldsymbol{Y}^i$. Note that we require that this random process is not influenced by controls. We propose to replace Problem (6) by:

$$\min_{\boldsymbol{X}^i, \boldsymbol{U}^i} \mathbb{E}\left(\sum_{t=0}^{T-1} \left(C_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t\right) + \mathbb{E}\left(\boldsymbol{\lambda}_t \mid \boldsymbol{Y}_t^i\right)^\top g_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t\right)\right) + K^i\left(\boldsymbol{X}_T^i\right)\right), \tag{7}$$

subject to constraints (6b)–(6f).

Let us first examine the special situation in which the information variable $\boldsymbol{Y}_t^i$ only depends on the current noise $\boldsymbol{W}_t$. The process $\boldsymbol{Y}^i$ does not add memory in the system so that Problem (7) can be solved using the standard DP equation:

$$V_T^i(x) = K^i(x), \qquad \forall x \in \mathbb{X}_T^i,$$

$$V_t^i(x) = \mathbb{E}\Big(\min_{u \in \mathbb{U}^i} C_t^i(x, u, \boldsymbol{W}_t) + \mathbb{E}\left(\boldsymbol{\lambda}_t \mid \boldsymbol{Y}_t^i\right)^\top g_t^i(x, u, \boldsymbol{W}_t) + V_{t+1}^i\left(f_t^i(x, u, \boldsymbol{W}_t)\right)\Big), \qquad \forall x \in \mathbb{X}_t^i.$$

The expectation quadrature only involves the noise variable $\boldsymbol{W}_t$. Remember, as explained in Remark 2, that we are in the "hazard-decision" setting: even though the control at each instant t depends on both $\boldsymbol{X}_t^i$ and $\boldsymbol{W}_t$, the Bellman function only depends on $\boldsymbol{X}_t^i$.

Because of the overtime independence of the information variables $\boldsymbol{Y}_t^i$, we have to solve DP equations whose dimension is the subsystem dimension n_i. Let us give three examples of choices for $\boldsymbol{Y}_t^i$.

Example 1 (Maximal information). One can choose to include in $\boldsymbol{Y}_t^i$ all the noise at time t. As already explained in Remark 3, the cost function and dynamics of a subsystem may only depend on a part of the whole noise $\boldsymbol{W}_t$ (a kind of *local* information denoted by $\boldsymbol{W}_t^i$ in Remark 3). Yet some *global* noise, denoted by $\boldsymbol{D}_t$ in Remark 3 may appear in the coupling constraint (e.g. a global power demand). Hence this maximal choice for the information variable makes the multiplier depend on both local and global information: this shall improve the subsystem's vision of the rest of the system and hence improves strategies. Note, however, that including all the noise at time t in the information variable is only possible in practice when the noise dimension is not too large. Indeed, the information variable appears in a conditional expectation, whose computation is subject to the curse of dimensionality.

Example 2 (Minimal information). On the opposite, one can choose $\boldsymbol{Y}_t^i = 0$ or any other constant. The dual stochastic process is then approximated by its expectation at every instant. Compared to the previous example, there is no conditional

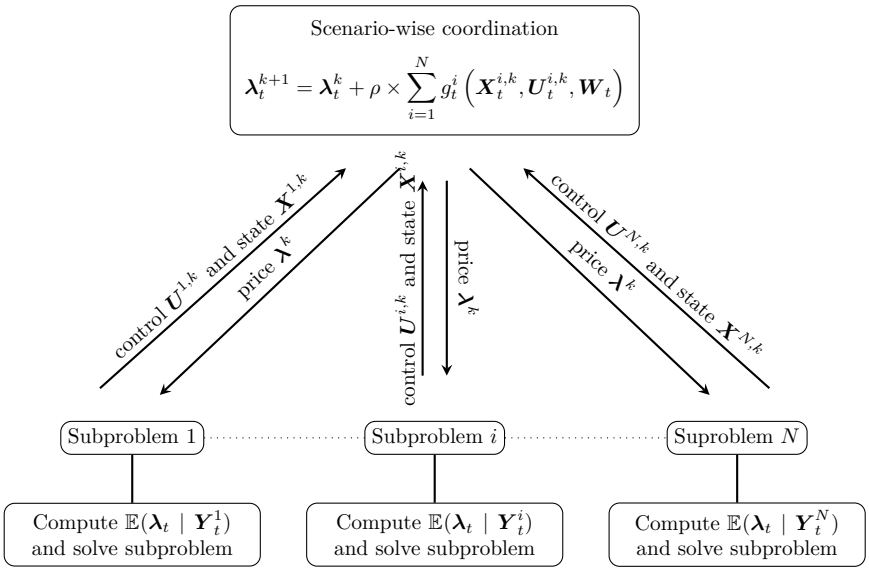

FIGURE 1. Dual Approximate Dynamic Programming

expectation anymore but one obtains a strategy that corresponds to the vision of an average price.

Example 3 (In between). One can choose $\boldsymbol{Y}_t^i$ of the form $h_t^i(\boldsymbol{W}_t)$. In practice, this choice will be guided by the intuition one has on which information mostly "explains" the optimal price of the system. One has to make a compromise between sufficient information to take reasonable actions and a not too large information variable to be able to compute the conditional expectation in (7).

Let us move towards the general case where one can choose to keep some information in memory. In other words, one can choose an information variable that has a Markovian dynamics, i.e. of the form $\boldsymbol{Y}_{t+1}^i = h_t^i(\boldsymbol{Y}_t^i, \boldsymbol{W}_{t+1})$. In order to derive a DP equation in this case, one has to augment the state vector by embedding $\boldsymbol{Y}_t^i$, that is the necessary memory to compute the next information variable. Thus, the Bellman function associated with the i-th subproblem depends, at time t, on both $\boldsymbol{X}_t^i$ and $\boldsymbol{Y}_{t-1}^i$. The DP equation writes:

$$V_t^i(x,y) = \mathbb{E}\Big(\min_u \quad C_t^i(x,u,\boldsymbol{W}_t) + \mathbb{E}\left(\boldsymbol{\lambda}_t^\top \;\middle|\; \boldsymbol{Y}_t^i\right) \cdot g_t^i(x,u,\boldsymbol{W}_t) + V_{t+1}^i\left(f_t^i(x,u,\boldsymbol{W}_t), \boldsymbol{Y}_t^i\right)\Big),$$
$$\text{with} \quad \boldsymbol{Y}_t^i = h_{t-1}^i(y, \boldsymbol{W}_t).$$

When solving this equation, one obtains controls as feedback functions on the local stock $\boldsymbol{X}_t^i$, the current noise $\boldsymbol{W}_t$ and the information variable $\boldsymbol{Y}_{t-1}^i$ of the previous time step. The index gap between information and stock variables comes from the "hazard-decision" setting: at time t, the information that is used to take decisions is the conjunction of the information kept in memory (that has index $t-1$) and of the noise observed at the current time step $\boldsymbol{W}_t$. The sketch of the DADP algorithm is depicted in Figure 1.

Example 4 (Perfect memory). The choice $\boldsymbol{Y}_t^i = (\boldsymbol{W}_0, \dots, \boldsymbol{W}_t)$ stands in the Markovian case. We have then $\mathbb{E}\left(\boldsymbol{\lambda}_t \;\middle|\; \boldsymbol{Y}_t^i\right) = \boldsymbol{\lambda}_t$. This choice hence allows us to model the dual variable perfectly, but the induced DP equation is unsolvable in practice.

Example 5 (Strugarek, 2006). In his PhD thesis, Strugarek exhibited a case when an exact model for the dual process can be obtained. His example is inspired from the kind of power management problem mentioned in the introduction, where N water reservoirs have to contribute to a global power demand, the rest of this demand being produced by fossil fuel. The noise at each time step t is composed of a scalar inflow $\boldsymbol{A}_t^i$ for each reservoir $i = 1, \dots, N$, and of a scalar power demand $\boldsymbol{D}_t$. The problem reads:

$$\min_{\boldsymbol{X},\boldsymbol{U}} \quad \mathbb{E}\left(\sum_{t=1}^{T-1}\sum_{j=1}^{n} c_j \frac{\left(\boldsymbol{U}_t^j\right)^2}{2} + \frac{\gamma_j}{2}\left(\boldsymbol{X}_t^j - x_1^j\right)^2\right), \tag{8a}$$

where c_j, $j = 1, \dots, N$ and γ_j, $j = 1, \dots, N$ are given real values, subject to dynamic constraints on reservoirs:

$$\boldsymbol{X}_{t+1}^j = \boldsymbol{X}_t^j + \boldsymbol{A}_{t+1}^j - \boldsymbol{U}_t^j, \qquad \forall t = 1, \dots, T-1, \forall j = 1, \dots, n, \tag{8b}$$

the power demand constraint:

$$\sum_{j=1}^{n} \boldsymbol{U}_t^j = \boldsymbol{D}_t, \qquad \forall t = 1, \dots, T-1, \tag{8c}$$

and the non-anticipativity constraint:

$$\boldsymbol{U}_t \text{ is } \sigma\left\{\boldsymbol{D}_s, s \leq t \,;\, \boldsymbol{A}_s, s \leq t\right\}\text{-measurable.} \tag{8d}$$

Let us denote $\boldsymbol{A}_t^\sigma := \sum_{i=1}^N \boldsymbol{A}_t^i$. The author then shows the following result.

Proposition 1 (Strugarek, 2006, Chapter V). *If random variables $(\boldsymbol{D}_t, \boldsymbol{A}_t)_{t=1,\dots,T}$ are independent over time, and if there exists $\alpha > 0$ such that $\gamma_j = \alpha c_j$, for all $j = 1, \dots, n$, then the optimal multiplier $\boldsymbol{\lambda}$ associated with the coupling constraints* (8c) *satisfies the following dynamics:*

$$\boldsymbol{\lambda}_1 = \frac{1}{\sum_{j=1}^n \frac{1}{c_j}}\left(\boldsymbol{D}_1(1-\alpha) - \alpha\sum_{s=2}^{T}\mathbb{E}\left(\boldsymbol{A}_s^\sigma\right) - \alpha\sum_{s=2}^{T-1}\mathbb{E}\left(\boldsymbol{D}_s\right)\right),$$

$$\boldsymbol{\lambda}_{t+1} = \boldsymbol{\lambda}_t + \frac{1}{\sum_{i=1}^n \frac{1}{c_i}}\Big[\boldsymbol{D}_{t+1}(1+\alpha) - \boldsymbol{D}_t - \alpha\mathbb{E}\left(\boldsymbol{D}_{t+1}\right) - \alpha\left(\boldsymbol{A}_{t+1}^\sigma - \mathbb{E}\left(\boldsymbol{A}_{t+1}^\sigma\right)\right)\Big], \qquad \forall t = 1, \dots, T-2.$$

This allows the solving of subproblems using DP in dimension 3. Note that this example enters our approach if one chooses $(\boldsymbol{Y}_t, \boldsymbol{D}_t)$ as an information variable, with:

$$\boldsymbol{Y}_1 = \frac{1}{\sum_{i=1}^n \frac{1}{c_i}}\left(\boldsymbol{D}_1(1-\alpha) - \alpha\sum_{s=2}^{T}\mathbb{E}\left(\boldsymbol{A}_s^\sigma\right) - \alpha\sum_{s=2}^{T-1}\mathbb{E}\left(\boldsymbol{D}_s\right)\right),$$

and, for all $t = 1, \dots, T-2$:

$$\boldsymbol{Y}_{t+1} = \boldsymbol{Y}_t + \frac{1}{\sum_{i=1}^n \frac{1}{c_i}}\Big[\boldsymbol{D}_{t+1}(1+\alpha) - \boldsymbol{D}_t - \alpha\mathbb{E}\left(\boldsymbol{D}_{t+1}\right) - \alpha\left(\boldsymbol{A}_{t+1}^\sigma - \mathbb{E}\left(\boldsymbol{A}_{t+1}^\sigma\right)\right)\Big].$$

We get back to the particular case when $\mathbb{E}(\boldsymbol{\lambda}_t \mid \boldsymbol{Y}_t^i) = \boldsymbol{\lambda}_t$, with a small dimensional information variable $\boldsymbol{Y}_t^i$. Note however that conditions of Proposition 1, especially the proportionality relation on costs, make little sense in practice.

2.2. **Convergence.** We now give convergence results about DADP and explain in more details the relation between the strategies it builds and the solution of the original problem (1). To make the paper self-contained, we recall in Appendix A the general results concerning duality in optimization, of which the properties of DADP are direct consequences.

The approximation made on the dual process gives us a tractable way of computing strategies for each one of the subsystems. Depending on the choice we make for the information variable, it is quite clear that some strategies will lead to better results than others, concerning the value of the dual problem or the satisfaction of the coupling constraint. Let us here state more precisely these facts.

From now on, we consider a unique information variable for all subsystems. We denote it by $\boldsymbol{Y}_t$ and define Hilbert spaces

$$\mathcal{Y}_t := \{\boldsymbol{\lambda}_t \in L^2(\Omega, \mathcal{A}, \mathbb{P}) : \boldsymbol{\lambda}_t \text{ is } \boldsymbol{Y}_t\text{-measurable}\},$$

for every $t = 0, \dots, T-1$.

Proposition 2. *Consider the following optimization problem:*

$$\min_{\boldsymbol{X},\boldsymbol{U}} \quad \mathbb{E}\left(\sum_{t=0}^{T-1}\sum_{i=1}^{N} C_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t\right) + \sum_{i=1}^{N} K^i\left(\boldsymbol{X}_T^i\right)\right), \tag{9a}$$

subject to the same constraints as in Problem (3) *except the coupling constraint* (3g) *which is replaced by:*

$$\mathbb{E}\left(\sum_{i=1}^{N} g_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t\right) \,\middle|\, \boldsymbol{Y}_t\right) = 0, \qquad \forall t = 0, \dots, T, . \tag{9b}$$

Suppose the Lagrangian associated with Problem (9) *has a saddle point. Then DADP solves Problem* (9).

Proof. The DADP algorithm consists in:

- given a price process, solving subproblems using the projection of this price process on $\mathcal{Y}_0 \times \cdots \times \mathcal{Y}_{T-1}$;
- updating the price process using a gradient formula.

Alternatively, one may consider that the gradient formula is composed with the projection operation in the updating formula. Therefore, this algorithm may also be viewed as a projected gradient algorithm which exactly solves the following max-min problem :

(10a)
$$\max_{\boldsymbol{\lambda}} \min_{\boldsymbol{X},\boldsymbol{U}} \quad \mathbb{E}\left(\sum_{t=0}^{T}\sum_{i=1}^{N}\left(C_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t\right) + \boldsymbol{\lambda}_t^\top g_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t\right)\right) + \sum_{i=1}^{N} K^i\left(\boldsymbol{X}_T^i\right)\right),$$

(10b)
$$\text{s.t.} \quad \boldsymbol{X}_{t+1}^i = f_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t\right), \qquad \forall t = 0, \dots, T-1, \forall i = 1, \dots, N,$$

(10c)
$$\boldsymbol{X}_0 = \boldsymbol{W}_0,$$

(10d)
$$\underline{x}_t^i \le \boldsymbol{X}_t^i \le \overline{x}_t^i, \qquad \forall t = 1, \dots, T, \forall i = 1, \dots, N,$$

(10e)
$$\underline{u}_t^i \le \boldsymbol{U}_t^i \le \overline{u}_t^i, \qquad \forall t = 0, \dots, T-1, \forall i = 1, \dots, N,$$

(10f)
$$\boldsymbol{U}_t \text{ is } \mathcal{A}_t\text{-measurable}, \qquad \forall t = 0, \dots, T,$$

(10g)
$$\boldsymbol{\lambda}_t \text{ is } \boldsymbol{Y}_t\text{-measurable}, \qquad \forall t = 0, \dots, T.$$

Observe that the max operation is restricted to a linear subspace defined by (10g).

Now, if within the inner product $\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \mathbb{E}\left(\boldsymbol{a}^\top \boldsymbol{b}\right)$, the variable $\boldsymbol{a}$ belongs to a given subspace, then the component of $\boldsymbol{b}$ which is orthogonal to that subspace

yields 0 in the inner product. Hence it is useless. Put in our context, the multiplier $\boldsymbol{\lambda}_t$ can only control the part of $g_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t\right)$ which has the same measurability as $\boldsymbol{\lambda}_t$. Thus, assuming the existence of a saddle point, that is, the max and min operations can be interchanged in Problem (10), this problem appears as the dual counterpart of Problem (9). □

Loosely speaking, DADP somehow consists in replacing an almost-sure constraint by a constraint involving a conditional expectation with respect to a so-called information variable. So it is once again clear that if we choose the information variable $\boldsymbol{Y}_t$ to be the whole history of the system, then we come back to the initial constraint and we in fact solve the original problem. This is the case of Example 4. On the contrary, putting no information at all in $\boldsymbol{Y}_t$ is the same as satisfying the coupling constraint only in expectation. This is the case of Example 2. Note however that it is generally a poor way of representing an almost-sure constraint.

The main difficulty is to find the information variable $\boldsymbol{Y}_t$ that is going to satisfy the coupling constraint in a fairly good way while keeping the solving process of the subproblems tractable.

We now state the convergence of the DADP algorithm. Let us introduce the objective function $J : \mathcal{U}_0 \times \cdots \times \mathcal{U}_{T-1} \to \mathbb{R}$ associated with strategy $\boldsymbol{U}$, i.e.:

$$\begin{aligned} J : \boldsymbol{U} \mapsto \quad & \mathbb{E}\left(\sum_{t=0}^{T-1}\sum_{i=1}^{N} C_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t\right) + \sum_{i=1}^{N} K^i\left(\boldsymbol{X}_T^i\right)\right), \\ & \text{with: } \boldsymbol{X}_0 = \boldsymbol{W}_0, \\ & \text{and: } \boldsymbol{X}_{t+1}^i = f_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t\right), \qquad \forall t = 0, \ldots, T-1, \forall i = 1, \ldots, N. \end{aligned}$$

Proposition 3. *If:*

(1) *J is convex, lower semi-continuous, Gâteaux differentiable,*
(2) *J is α-strongly convex,*
(3) *all g_t^i are linear and c-Lipschitz continuous,*
(4) *the Lagrangian associated with Problem (9) has a saddle point $(\overline{\boldsymbol{U}}, \overline{\boldsymbol{\lambda}})$,*
(5) *the step-size ρ of the algorithm is such that $0 < \rho < 2\frac{\alpha}{c^2}$,*

Then:

(1) *there exists a unique solution $\overline{\boldsymbol{U}}$ of Problem (9),*
(2) *DADP converges in the sense that :*

$$\boldsymbol{U}^k \underset{k\to+\infty}{\longrightarrow} \overline{\boldsymbol{U}} \text{ in } \mathcal{U}_0 \times \cdots \times \mathcal{U}_{T-1},$$

(3) *the sequence $(\boldsymbol{\lambda}^k)_{k\geq 0}$ is bounded and every cluster point $\overline{\boldsymbol{\lambda}}$ in the weak topology is such that $(\overline{\boldsymbol{U}}, \overline{\boldsymbol{\lambda}})$ is a saddle point of the Lagrangian associated with Problem (9).*

Proof. The convergence of the algorithm is then a direct application of Theorem 1, Appendix A. □

Note that assumptions of Proposition 3 plus the qualification of constraint (9b) ensure that the Lagrangian associated with Problem (9) has a saddle point.

3. Numerical experiment

We now show the efficiency of DADP on two numerical examples. The first one comes from a previous paper (Barty et al., 2010) in which the authors developped a preliminary version of DADP (see §1.4). We show in §3.2 the good performance of the new version of DADP. The second one, in §3.3, is an application to a more realistic power management problem.

3.1. **Computing conditional expectations.** Within the DADP procedure, at each iteration, we have to compute conditional expectations in the criteria (7) of the subproblems. In order to compute these conditional expectations, we used Generalized Additive Models (GAMs), that were introduced by Hastie and Tibshirani (1990). The estimate takes the form:

$$\mathbb{E}\left(\boldsymbol{Z} \mid \boldsymbol{P}_1, \ldots, \boldsymbol{P}_n\right) \simeq \sum_{i=1}^{n} f_i\left(\boldsymbol{P}_i\right).$$

Functions f_i are splines (piecewise polynoms) whose characteristics are optimized by cross-validation on the input statistical data. Our purpose here is not to explain in details this methodology. The interested reader will find further explanations about this model and its implementation in the book by Wood (2006). We used an easy-to-use implementation that is available within the free statistical software R (R Development Core Team, 2009). The GAM toolkit, called *mgcv*, also returns useful indicators concerning the quality of the estimation. In particular, we use the deviance indicator, which takes value 0 if $\boldsymbol{Z}$ is estimated as poorly as by its expectation $\mathbb{E}\left(\boldsymbol{Z}\right)$ and value 1 if the estimate is exact, i.e. if $\sum_{i=1}^{n} f_i(\boldsymbol{P}_i) = \boldsymbol{Z}$.

Remark 4 (Kernel estimator). We chose to use GAMs to compute conditional expectations after a numerical comparison with the more classical kernel regression methods (Nadaraya, 1964, Watson, 1964) also available in the R environment. Even though both of them gave similar results, GAMs appeared to be several times faster than the kernel method on our problem.

3.2. **Back to an example from a previous paper.** We first implement the new version of DADP algorithm on a simple power management problem introduced by Barty et al. (2010). On this small-scale example, we are able to compare DADP results to those obtained by DP and to illustrate the theoretical results described above. Let us first recall this example. Consider a power producer who owns two types of power plants:

- Two hydraulic plants that are characterized at each time step t by their water stock $\boldsymbol{X}_t^i$ and power production $\boldsymbol{U}_t^i$, and receive water inflows $\boldsymbol{A}_{t+1}^i$, $i = 1, 2$. Such units are usually cost-free. We however impose small quadratic costs on the hydraulic power productions in order to ensure strong convexity.
- One thermal unit with a production cost that is quadratic with respect to its production $\boldsymbol{U}_t^3$. There are no dynamics associated with this unit.

Using these plants, the power producer must supply a power demand $\boldsymbol{D}_t$ at each time step t, over a discrete time horizon of $T = 25$ time steps. All noises, i.e. demand $\boldsymbol{D}_t$ and inflows $\boldsymbol{A}_t^1$ and $\boldsymbol{A}_t^2$ are supposed to be overtime independent noise processes. The interested reader may find more details on this numerical experiment in the previous paper by Barty et al. (2010).

The problem reads:

$$\min_{\boldsymbol{X},\boldsymbol{U}} \quad \mathbb{E}\left(\sum_{t=0}^{T-1}\left(\epsilon\left(\boldsymbol{U}_t^1\right)^2+\epsilon\left(\boldsymbol{U}_t^2\right)^2+L_t\left(\boldsymbol{U}_t^3\right)\right)+K^1\left(\boldsymbol{X}_T^1\right)+K^2\left(\boldsymbol{X}_T^2\right)\right) \tag{11a}$$

$$\text{s.t.} \quad \boldsymbol{X}_{t+1}^i=\boldsymbol{X}_t^i-\boldsymbol{U}_t^i+\boldsymbol{A}_{t+1}^i, \qquad \forall i=1,2, \quad \forall t=0,\dots,T-1, \tag{11b}$$

$$\boldsymbol{U}_t^1+\boldsymbol{U}_t^2+\boldsymbol{U}_t^3=\boldsymbol{D}_t, \qquad \forall t=0,\dots,T-1, \tag{11c}$$

$$\underline{x}^i\le\boldsymbol{X}_t^i\le\overline{x}^i, \qquad \forall i=1,2, \quad \forall t=1,\dots,T, \tag{11d}$$

$$0\le\boldsymbol{U}_t^i\le\overline{u}^i, \qquad \forall i=1,2, \quad \forall t=0,\dots,T-1, \tag{11e}$$

$$0\le\boldsymbol{U}_t^3, \qquad \forall t=0,\dots,T-1, \tag{11f}$$

$$\boldsymbol{U}_t^i \text{ is } \sigma\left\{\boldsymbol{D}_0,\boldsymbol{A}_0^1,\boldsymbol{A}_0^2,\dots,\boldsymbol{D}_t,\boldsymbol{A}_t^1,\boldsymbol{A}_t^2\right\}\text{-measurable}, \quad \forall i=1,2,3. \tag{11g}$$

In this problem, the state $\boldsymbol{X}_t$ is two-dimensional, hence DP remains numerically tractable and we can use the DP solution as a reference. In order to use DADP, we choose an information variable $\boldsymbol{Y}_t$ at time t that is equal to the power demand $\boldsymbol{D}_t$. This comes from the insight that the power demand is a "global" information and has all reasons to be useful to the subproblems.

Remark 5 (Primal feasibility). In order to validate the method, it has to be evaluated within a simulation procedure. For the evaluation to be fair, the strategy must be feasible. Yet, as explained in §2.2, DADP does not ensure that the coupling constraint (3f) is satisfied. To circumvent this difficulty, the thermal unit strategy is chosen in the simulation process so as to ensure feasibility of the coupling constraint, i.e.:

$$\boldsymbol{U}_t^3=\boldsymbol{D}_t-\left(\boldsymbol{U}_t^1+\boldsymbol{U}_t^2\right). \tag{12}$$

That is, DADP returns three strategies, for each of the hydraulic units and for the thermal unit. However, we use relation (12) for the thermal strategy during simulations in order to ensure demand satisfaction and give an estimation of the cost of the DADP strategy.

We run the algorithm for 20 iterations and depict its behaviour in Figure 2. We draw the dual cost (evaluation of the dual function with the current strategy) and the primal cost (the one with all constraints satisfied) at each iteration. Each point of the primal and dual curves is computed by Monte Carlo simulation over 500 scenarios. We observe the regular increase of the dual function, as expected, and the decrease of the primal function. The distance between the primal and dual costs is an upper bound for the distance to the optimal value that graphically, in this case, seems quite tight.

Moreover, the GAM toolkit used to compute the conditional expectations of the form $\mathbb{E}\left(\boldsymbol{\lambda}_t \mid \boldsymbol{D}_t\right)$ returns that the deviance, i.e. the quality of the explanation of $\boldsymbol{\lambda}_t$ by $\boldsymbol{D}_t$ is 98.5%. This indicates that the marginal cost of the system is almost perfectly explained by the time variable and the power demand. Otherwise stated, using $\mathbb{E}\left(\boldsymbol{\lambda}_t \mid \boldsymbol{D}_t\right)$ instead of using $\boldsymbol{\lambda}_t$ within Problem (11) does not alter too much the quality of the solution.

3.3. **A larger-scale SOC problem.** We now apply DADP on a real-life power management problem, inspired by a case encountered at EDF, which is the major European power producer. We do not give the exact order of magnitude for costs and productions because of confidentiality issues. We consider :

- a power demand on a single node (we neglect network issues) at each instant of a finite time horizon of 163 weeks (one time step per week);
- 7 (hydraulic) stocks which are in fact aggregations of many smaller stocks;

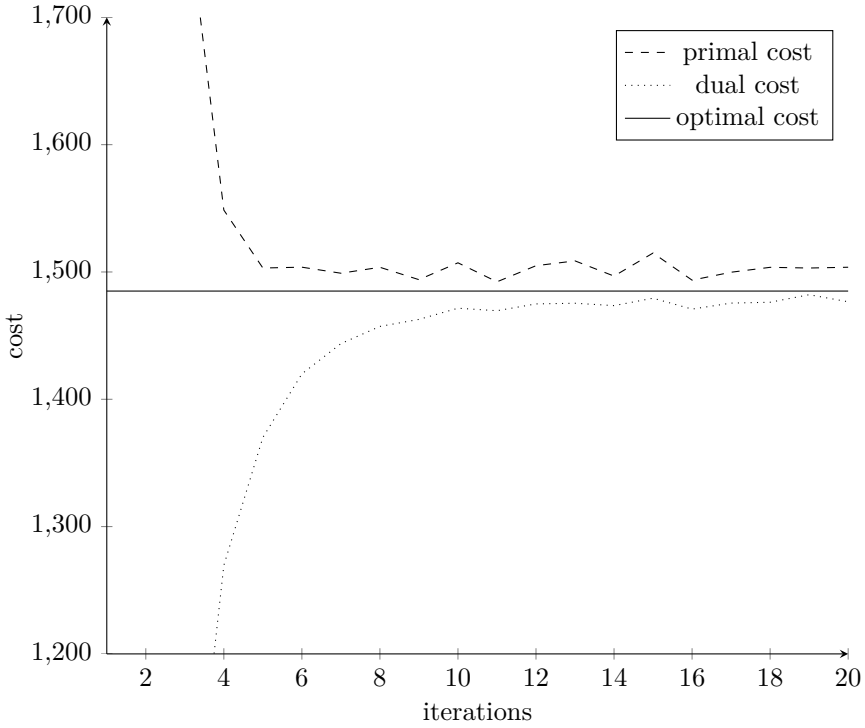

FIGURE 2. Primal, dual and optimal costs with respect to the number of iterations

- 122 other (thermal) power units with no stock constraints.

All the thermal power units are aggregated so that the thermal cost $\boldsymbol{C}_t$ at each time t only depends on the total thermal production $\boldsymbol{U}_t^{\mathrm{th}}$ and forms a quadratic cost. We note $\boldsymbol{C}_t$ using bold letters, which means that this thermal cost is random, because of the breakdowns that may happen on thermal power plants.

The problem reads:

(13a)
$$\min_{\boldsymbol{X},\boldsymbol{U}} \quad \mathbb{E}\left(\sum_{t=0}^{T-1} \boldsymbol{C}_t\left(\boldsymbol{U}_t^{\mathrm{th}}\right)\right),$$

subject to hydraulic stock dynamics :

(13b) $$\boldsymbol{X}_0^i = x_0^i, \qquad \forall i = 1,\dots,7,$$

(13c) $$\boldsymbol{X}_{t+1}^i = \boldsymbol{X}_t^i - \boldsymbol{U}_t^i + \boldsymbol{A}_t^i, \qquad \forall i = 1,\dots,7, \forall t = 0,\dots,T-1,$$

power demand constraints :

(13d) $$\sum_{i=1}^{7} \boldsymbol{U}_t^i + \boldsymbol{U}_t^{\mathrm{th}} = \boldsymbol{D}_t, \qquad \forall t = 0,\dots,T-1,$$

bound constraints on stocks and controls :

$$\underline{u}_t^{\text{th}} \leq \boldsymbol{U}_t^{\text{th}} \leq \overline{u}_t^{\text{th}}, \qquad \forall t = 0, \dots, T-1, \tag{13e}$$

$$\underline{u}_t^i \leq \boldsymbol{U}_t^i \leq \overline{u}_t^i, \qquad \forall i = 1, \dots, 7, \forall t = 0, \dots, T-1, \tag{13f}$$

$$\underline{x}_t^i \leq \boldsymbol{X}_t^i \leq \overline{x}_t^i, \qquad \forall i = 1, \dots, 7, \forall t = 0, \dots, T, \tag{13g}$$

and non-anticipativity constraints :

$$\boldsymbol{U}_t^i \text{ is } (\boldsymbol{W}_0, \dots, \boldsymbol{W}_t)\text{-measurable}, \qquad \forall i = 1, \dots, 7, \forall t = 0, \dots, T-1, \tag{13h}$$

$$\boldsymbol{U}_t^{\text{th}} \text{ is } (\boldsymbol{W}_0, \dots, \boldsymbol{W}_t)\text{-measurable}, \qquad \forall t = 0, \dots, T-1, \tag{13i}$$

with $\boldsymbol{W}_t := (\boldsymbol{A}_t, \boldsymbol{C}_t, \boldsymbol{D}_t)$ being the set of all noises that affect the system at time t.

Because we consider 7 stocks, we are unable to use DP directly on this problem. In order to obtain a reference point, we use an aggregation method introduced by Turgeon (1980) and currently in use at EDF. This numerical method is known to be especially well-suited for the problem under consideration. It consists in solving N subproblems (7 in our case) by 2-dimensional DP, each subproblem relying on a particular power unit, instead of one N-dimensional DP problem. The idea is, for every unit, to look for strategies that depend on the stock of the unit and on an aggregation of the remaining stocks.

We then make use of DADP using three different choices for the information variable $\boldsymbol{Y}_t$.

- In the first setting, we replace the price at each time step by its expectation. In other words, we explain the price only by the time variable t. According to Proposition 3, we are in fact solving Problem (13) with constraint (13d) replaced by its expectation. Then we are able to solve each subproblem i by DP in dimension 1 (the stock variable of unit i) and we obtain strategies that depend, for each unit i and each instant t, on the stock $\boldsymbol{X}_t^i$ and the inflow $\boldsymbol{A}_t^i$.
- In the second setting, we replace the price at each time step by its conditional expectation with respect to the power demand. Put differently, we explain the price by time and demand. We still have to solve a 1-dimensional DP equation and we obtain for each instant t a strategy that depends on $\boldsymbol{X}_t^i$, $\boldsymbol{A}_t^i$ and $\boldsymbol{D}_t$.
- In the third setting, we replace the price at each instant by its conditional expectation with respect to the power demand and the thermal availability[5] $\overline{\boldsymbol{P}}_t$. We then obtain a strategy that depends, for every unit i and every instant t, on $\boldsymbol{X}_t^i$, $\boldsymbol{A}_t^i$, $\boldsymbol{D}_t$ and $\overline{\boldsymbol{P}}_t$.

The behaviour of the algorithm in the second setting is depicted in Figure 3. We observe the increase of the dual value and the decrease of the primal value, the latter value stabilizing rapidly to a value close to the one of the aggregation method. Even though we are aware that only 10 iterations is generally much too less for this kind of primal-dual algorithm, it seems like the primal cost does not evolve significantly after 10 iterations.

In order to compare the three settings, we simulate the corresponding strategies[6] on a large set of i.i.d. noise scenarios and compute both the mean cost and confidence interval for each strategy. The results are presented in Table 1. The "Deviance" column gives the deviance indicator returned by the GAM procedure

[5]The thermal availability is a scalar variable computed out of the thermal cost function $\boldsymbol{C}_t$. It gives insight on how tense the thermal generation mix is.

[6]As in the previous example, the thermal unit strategy is chosen so as to ensure feasibility of the coupling constraint (see Remark 5).

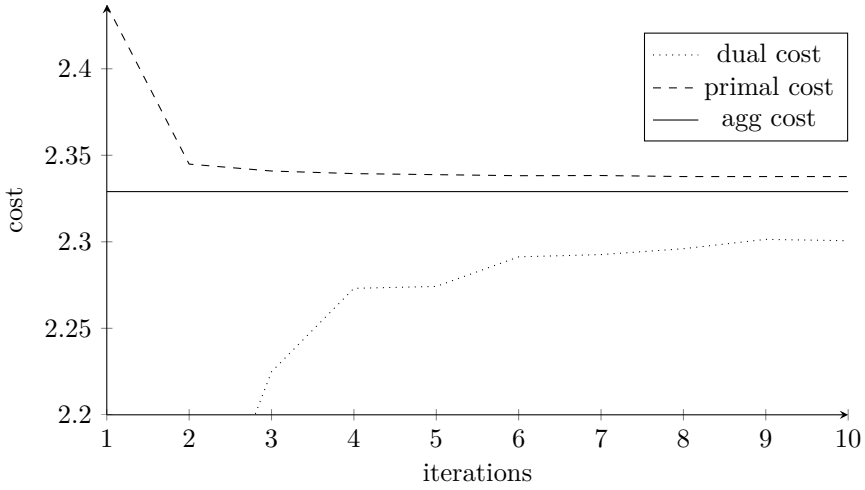

FIGURE 3. Primal and dual costs along with iterations compared to the aggregation method

	Mean cost	$CI_{95\%}$	Deviance
First setting	2.363	$1.3 \cdot 10^{-2}$	50.0%
Second setting	2.340	$1.3 \cdot 10^{-2}$	82.4%
Third setting	2.338	$1.3 \cdot 10^{-2}$	86.1%

TABLE 1. Results for DADP

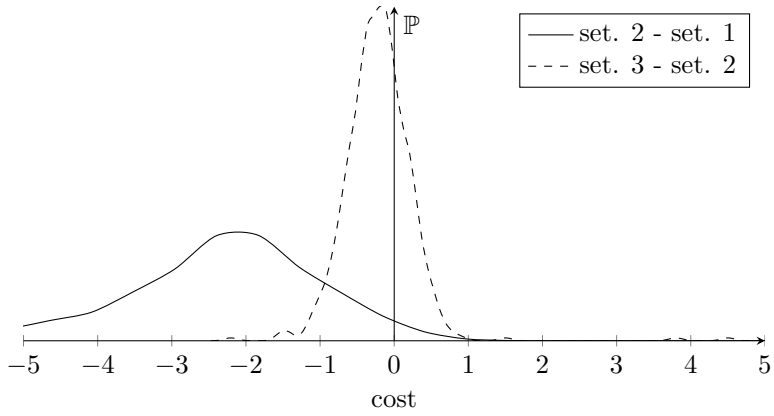

FIGURE 4. Distribution of cost differences between settings of DADP

for the estimation of the conditional expectation of the price with respect to the information variable. We observe that the DADP strategy still benefits from a good choice for the information variable $\boldsymbol{Y}_t$: it appears from the mean costs comparison that adding information within the estimator improves the quality of the estimation. The mean costs differences are however not so easy to compare for the two last experiments, because the confidence interval is too large compared to the cost values. Thus we compute for each scenario the gap between costs obtained by two different strategies and draw in Figure 4 the associated probability distributions. It becomes clearer that adding the thermal availability in the information variable

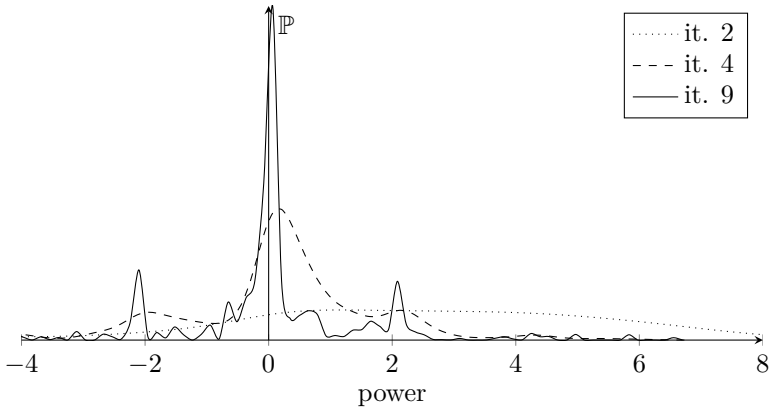

FIGURE 5. Distribution of the production/demand gap for a given time step

improves the strategy: the major part of the probability weight when comparing settings 2 and 3 is negative.

As a last point, let us numerically verify that Proposition 3 holds in our example, for instance in the first setting. Remember that, in this case, our algorithm aims at satisfying the coupling constraint only in expectation. We draw in Figure 5 the probability distribution of the production/demand gap at several iterations. We observe that, along with iterations, the distribution of this gap becomes symmetric with respect to 0, the corresponding expectation hence being equal to zero.

CONCLUSION

We presented an original algorithm for solving a certain kind of large-scale stochastic optimal control problems. It is based on an approximate Lagrangian decomposition: the Lagrange multiplier, which is a stochastic process in this context, is projected using a conditional expectation with respect to another stochastic process called the information process. This information process is chosen a priori and, when it has a limited memory, the solving of subproblems becomes tractable. We give theoretical results concerning the convergence of the algorithm and show how it actually solves an approximate problem, whose relation with the original problem is driven by the choice of information variable. Finally, we show on two numerical examples the efficiency of the approach.

Future works will be concerned with the application of this algorithm to more general problem structures, like chained subsystems or networks.

APPENDIX A. DUALITY IN CONVEX OPTIMIZATION

The results presented here come from the paper by Cohen (1980a). Let $\mathcal{U}$ and Λ be Hilbert spaces[7], and $\mathcal{U}^{\mathrm{ad}}$ and Λ^{ad} be subsets of $\mathcal{U}$ and Λ (respectively). Moreover, let us define a function $L : \mathcal{U} \times \Lambda \to \mathbb{R}$. We describe here the relations that link the so-called primal problem:

$$\inf_{u \in \mathcal{U}^{\mathrm{ad}}} \sup_{\lambda \in \Lambda^{\mathrm{ad}}} L(u, \lambda), \tag{14}$$

[7]These results can be generalized to Banach spaces (see Ekeland and Temam, 1999), but this is not necessary for our purpose.

to its dual counterpart:

$$\sup_{\lambda\in\Lambda^{\mathrm{ad}}}\inf_{u\in\mathcal{U}^{\mathrm{ad}}} L\left(u,\lambda\right).$$

$\mathcal{U}$ is called the primal space while Λ is called the dual one.

Definition 1 (Saddle point)**.** A pair $(\overline{u},\overline{\lambda})\in\mathcal{U}^{\mathrm{ad}}\times\Lambda^{\mathrm{ad}}$ is called a saddle point of L on $\mathcal{U}^{\mathrm{ad}}\times\Lambda^{\mathrm{ad}}$ if:

$$L\left(\overline{u},\lambda\right)\leq L\left(\overline{u},\overline{\lambda}\right)\leq L\left(u,\overline{\lambda}\right),\qquad\forall u\in\mathcal{U}^{\mathrm{ad}},\forall\lambda\in\Lambda^{\mathrm{ad}}.$$

Let us now concentrate on the case where function L corresponds to the Lagrangian of an optimization problem:

$$L\left(u,\lambda\right)=J\left(u\right)+\left\langle\lambda,g\left(u\right)\right\rangle.$$

The Uzawa algorithm is defined as follows. Take an initial value $\lambda_0\in\Lambda^{\mathrm{ad}}$. At each iteration $n\geq 0$, compute u_n by minimizing $J\left(u\right)+\left\langle\lambda_n,g\left(u\right)\right\rangle$, and update λ_n using the following rule:

$$\lambda_{n+1}=\Pi_{\Lambda^{\mathrm{ad}}}\left(\lambda_n+\rho_n g\left(u_n\right)\right),$$

with ρ_n some positive value. The following theorem gives conditions for the sequence $(u_n)_{n\geq 0}$ to converge to the optimum of Problem (14).

Theorem 1 (Cohen, 1980a, Theorem 6.1)**.** *If:*

(1) *J is convex, lower semi-continuous, Gâteaux differentiable,*
(2) *J is α-strongly convex,*
(3) *g is linear and c-Lipschitz continuous,*
(4) *L has at least a saddle point $(\overline{u},\overline{\lambda})$,*
(5) *the step-size ρ of the algorithm is such that $0<\rho<2\frac{\alpha}{c^2}$,*

then:

(1) *$\overline{u}$ is unique and is a solution of Problem* (14)*,*
(2) *Uzawa's algorithm converges in the sense that :*

$$u_n\underset{n\to+\infty}{\longrightarrow}\overline{u}\ \text{in}\ \mathcal{U},$$

(3) *the sequence $(\lambda_n)_{n\geq 0}$ is bounded and every cluster point $\overline{\lambda}$ in the weak topology is such that $(\overline{u},\overline{\lambda})$ is a saddle point of L.*

Given the other assumptions of the theorem, assumption (4) is satisfied as long as the dualized constraint satisfies a so-called "qualification" condition. In addition, the latter is always satisfied for affine constraints, which is the case in our application.

Appendix B. A lemma about decomposition

We here depict in more details the reasons why a Stochastic Optimal Problem (SOC) involving N independent[8] subsystems is equivalent, under certain conditions, to N problems where each one involves only one of the subsystems. Though this result may seem trivial at first sight, it is not true in general: the interested reader will find a counter example in the paper by Cohen (1980b).

Lemma 1. *Consider the following problem:*

$$\min_{\boldsymbol{X},\boldsymbol{U}}\quad\mathbb{E}\left(\sum_{t=0}^{T-1}\sum_{i=1}^{N}C_t^i\left(\boldsymbol{X}_t^i,\boldsymbol{U}_t^i,\boldsymbol{W}_t^i,\boldsymbol{Z}_t\right)+\sum_{i=1}^{N}K^i\left(\boldsymbol{X}_T^i\right)\right)\tag{15a}$$

[8] in a sense that is made clear in Lemma 1

subject to dynamics constraints:

$$\boldsymbol{X}_{t+1}^i = f_t^i\left(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t^i, \boldsymbol{Z}_t\right), \qquad \forall t = 0, \dots, T-1, \forall i = 1, \dots, N, \tag{15b}$$

$$\boldsymbol{X}_0^i \text{ is given}, \qquad \forall i = 1, \dots, N, \tag{15c}$$

as well as bound constraints:

$$\underline{x}_t^i \leq \boldsymbol{X}_t^i \leq \overline{x}_t^i, \qquad \forall t = 0, \dots, T, \forall i = 1, \dots, N, \tag{15d}$$

$$\underline{u}_t^i \leq \boldsymbol{U}_t^i \leq \overline{u}_t^i, \qquad \forall t = 0, \dots, T-1, \forall i = 1, \dots, N, \tag{15e}$$

and the non-anticipativity constraint:

$$\boldsymbol{U}_t^i \text{ is } \mathcal{A}_t\text{-measurable}, \qquad \forall t = 0, \dots, T-1, \forall i = 1, \dots, N, \tag{15f}$$

where $\mathcal{A}_t$ is the σ-algebra generated by the random variables $\{\boldsymbol{W}_s^i, \boldsymbol{Z}_s\}$ for $i = 1, \dots, N$ and $s = 0, \dots, t$. We assume that:

- *the $\boldsymbol{W}_\cdot^i$'s and $\boldsymbol{Z}_\cdot$ are all white noise processes,*
- *that $\boldsymbol{W}_t^i$ is not necessarily independent from $\boldsymbol{W}_t^j$ for $j \neq i$ nor from $\boldsymbol{Z}_t$.*

Then, the optimal feedback solution is partially decentralized*, that is, each optimal decision $\boldsymbol{U}_t^i$, that may a priori depend on the whole $\boldsymbol{X}_t$ and the whole $\boldsymbol{W}_t$ and $\boldsymbol{Z}_t$ according to* (15f)*, indeed only depends on $(\boldsymbol{X}_t^i, \boldsymbol{W}_t^i, \boldsymbol{Z}_t)$; the Bellman function $V_t(\boldsymbol{X}_t)$ is additive ($V_t(\boldsymbol{X}_t) = \sum_{i=1}^N V_t^i(\boldsymbol{X}_t^i)$) and the optimal solution only involves the* marginal *probability laws of the pairs $(\boldsymbol{W}_t^i, \boldsymbol{Z}_t)$ but not the* joint *probability laws of the pairs $(\boldsymbol{W}_t, \boldsymbol{Z}_t)$.*

Proof. The proof is by induction over time. The statement that V is additive is true at the final time T since the final cost K is additive. Assume this is true from T to $t+1$ (backward). The Bellman equation at t reads:

$$V_t(x) = \mathbb{E}\left(\min_u \sum_{i=1}^N C_t^i(x^i, u^i, \boldsymbol{W}_t^i, \boldsymbol{Z}_t) + \sum_{i=1}^N V_{t+1}^i\big(f_t^i(x^i, u^i, \boldsymbol{W}_t^i, \boldsymbol{Z}_t)\big) \right),$$

in which

- the minimization operation is done over an expression is which x, $\boldsymbol{Z}_t$ and $\boldsymbol{W}_t^i$ are fixed (hazard-decision scheme) and the $\arg\min$ in u parametrically depends on those values (which yields the optimal feedback function) ;
- the minimization operation is subject to the bound constraints (15e) for u^i and (15d) for $f_t^i(x^i, u^i, \boldsymbol{W}_t^i, \boldsymbol{Z}_t)$;
- the expectation concerns random variables $(\boldsymbol{W}_t, \boldsymbol{Z}_t)$ whereas x is still fixed ($\boldsymbol{X}_t$ and $(\boldsymbol{W}_t, \boldsymbol{Z}_t)$ are independent from each other, thus this expectation may be considered as a conditional expectation knowing that $\boldsymbol{X}_t = x$): this yields a function of x, namely $V_t(\cdot)$.

Now observe that, at the minimization stage, each u^i is involved into a separate expression depending only on x^i, $\boldsymbol{W}_t^i$ and $\boldsymbol{Z}_t$ subject also to independent constraints, hence the claimed partially decentralized optimal feedback. Then, at the outer expectation stage, we get a sum of functions of x^i and $(\boldsymbol{W}_t^i, \boldsymbol{Z}_t)$: thus only the marginal probability law of each pair $(\boldsymbol{W}_t^i, \boldsymbol{Z}_t)$ is involved in the expectation of the corresponding term in this sum, and the result is an additive function of the x^i, which completes the proof by induction. □

Let us now comment some particular cases.

- If $\boldsymbol{Z}_t$ is absent and if $\boldsymbol{W}_\cdot^i$ and $\boldsymbol{W}_\cdot^j$ are independent whenever $j \neq i$, then the overall problem is obviously made up of N independent subproblems; the optimal feedbacks are fully decentralized (that is $\boldsymbol{U}^i$ is in closed loop

on $(\boldsymbol{X}^i, \boldsymbol{W}^i)$), and the optimal controls $\boldsymbol{U}_t^i$ and $\boldsymbol{U}_t^j$ are also independent random variables whenever $j \neq i$.

- If we drop the independency assumption about $\boldsymbol{W}_t^i$ and $\boldsymbol{W}_t^j$, then the same subproblems still provide the overall problem solution with decentralized feedbacks, but $\boldsymbol{U}_t^i$ and $\boldsymbol{U}_t^j$ are no longer independent.
- Another "extreme" situation is when only the "shared" noise $\boldsymbol{Z}$ is present in all subsystems (the $\boldsymbol{W}^i$'s are supposed absent for the sake of clarity but now, $\boldsymbol{Z}$ may be thought as the concatenation of all the $\boldsymbol{W}^i$'s). The conclusions of the lemma are of course still valid, that is, the Bellman function is still additive and each term of this sum can be calculated in a separate subproblem, yielding a feedback on $(\boldsymbol{X}^i, \boldsymbol{Z})$. However the price to be payed for the presence of this shared random variable is that, first, the minimization operation in the Bellman function is parametrized by both x^i and $\boldsymbol{Z}_t$, which may be costly if $\boldsymbol{Z}_t$ is of large dimension, and, second, the outer expectation in this Bellman equation involves a multiple integral over that vector $\boldsymbol{Z}_t$, which may also be costly.

References

L. Bacaud, C. Lemaréchal, A. Renaud, and C. A. Sagastizábal. Bundle methods in stochastic optimal power management: A disaggregated approach using preconditioner. *Computational Optimization and Applications*, 20(3):227–244, 2001.

K. Barty. *Contributions à la discrétisation des contraintes de mesurabilité pour les problèmes d'optimisation stochastique*. Thèse de doctorat, École Nationale des Ponts et Chaussées, 2004.

K. Barty, J.-S. Roy, and C. Strugarek. A stochastic gradient type algorithm for closed-loop problems. *Mathematical Programming, Series A*, 119(1):51–78, June 2009. doi: 10.1007/s10107-007-0201-x.

K. Barty, P. Carpentier, and P. Girardeau. Decomposition of large-scale stochastic optimal control problems. *RAIRO Operations Research*, 44(3):167–183, 7 2010. doi: 10.1051/ro/2010013.

R. Bellman. *Dynamic Programming*. Princeton University Press, New Jersey, 1957.

R. Bellman and S. E. Dreyfus. Functional approximations and dynamic programming. *Math tables and other aides to computation*, 13:247–251, 1959.

D. P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, 2 edition, 2000. ISBN 1886529094.

D. P. Bertsekas and J. N. Tsitsiklis. *Neuro-Dynamic Programming*. Athena Scientific, 1996.

B. Bouchard and X. Warin. Monte-Carlo Valorisation of American options: facts and new algorithms to improve existing methods. http://www.ceremade.dauphine.fr/~bouchard/pdf/BW10.pdf, 2010.

P. Carpentier, C. Cohen, J.-C. Culioli, and A. Renaud. Stochastic optimization of unit commitment: a new decomposition framework. *IEEE Transactions on Power Systems*, 11(2):1067–1073, 5 1996.

G. Cohen. Auxiliary Problem Principle and decomposition of optimization problems. *Journal of Optimization Theory and Applications*, 32(3):277–305, 11 1980a.

G. Cohen. Information Exchange Between Independent Stochastic Systems. *Journal of Optimization Theory and Applications*, 32(2):201–210, 10 1980b.

G. Cohen and J.-C. Culioli. Decomposition Coordination Algorithms for Stochastic Optimization. *SIAM J. Control Optimization*, 28(6):1372–1403, 1990.

D. P. de Farias and B. Van Roy. The Linear Programming Approach to Approximate Dynamic Programming. *Oper. Res.*, 51(6):850–856, 2003.

I. Ekeland and R. Temam. *Convex Analysis and Variational Problems*, volume 28 of *Classics in Applied Mathematics*. SIAM, 1999.

T. J. Hastie and R. J. Tibshirani. *Generalized Additive Models*. Chapman & Hall/CRC, 1990.

J. L. Higle and S. Sen. *Stochastic Decomposition: A Statistical Method for Large Scale Stochastic Linear Programming*. Kluwer Academic Publishers, Dordrecht, 1996.

F. A. Longstaff and E. S. Schwartz. Valuing american options by simulation: A simple least squares approach. *Review of Financial Studies*, 14(1):113–147, 2001.

E. A. Nadaraya. On estimating regression. *Theory of Probability and its applications*, 10:186–190, 1964.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2009. URL http://www.R-project.org. ISBN 3-900051-07-0.

A. Ruszczyński and A. Shapiro, editors. *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*. Elsevier, 2003.

A. Shapiro. On complexity of multistage stochastic programs. *Operations Research Letters*, 34:1–8, 2006.

A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming*. Society for Industrial and Applied Mathematics, Philadelphia, 2009.

C. Strugarek. *Approches variationnelles et autres contributions en optimisation stochastique*. PhD thesis, École Nationale des Ponts et Chaussées, 5 2006.

J. N. Tsitsiklis and B. Van Roy. Feature-based methods for large-scale dynamic programming. *Machine Learning*, 22:59–94, 1996.

A. Turgeon. Optimal operation of multi-reservoir power systems with stochastic inflows. *Water Resources Research*, 16(2):275–283, 1980.

P. Vezolle, S. Vialle, and X. Warin. Large Scale Experiment and Optimization of a Distributed Stochastic Control Algorithm. Application to Energy Management Problems. In *International workshop on Large-Scale Parallel Processing (LSPP 2009)*, Rome, Italy, 2009. ISBN 978-1-4244-3750-4.

G. S. Watson. Smooth regression analysis. *Shankya Series A*, 26:359–372, 1964.

S. N. Wood. *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC, 2006.

K. Barty, EDF R&D, 1 avenue du Général de Gaulle, F-92141 Clamart Cedex, France.
E-mail address: kengy.barty@edf.fr

P. Carpentier, ENSTA ParisTech, 32 boulevard Victor, 75739 Paris Cedex 15, France.
E-mail address: pierre.carpentier@ensta-paristech.fr

G. Cohen, Université Paris-Est, CERMICS, École des Ponts ParisTech, 6 & 8 avenue Blaise Pascal, 77455 Marne-la-Vallée Cedex 2.
E-mail address: guy.cohen@mail.enpc.fr

P. Girardeau, EDF R&D, 1 avenue du Général de Gaulle, F-92141 Clamart Cedex, France, also with Université Paris-Est, CERMICS and ENSTA ParisTech.
E-mail address: pierre.girardeau@cermics.enpc.fr